

INTERACTION EFFECTS IN MULTIELEMENT DESIGNS: INEVITABLE, DESIRABLE, AND IGNORABLE

ANN HIGGINS HAINS

UNIVERSITY OF WISCONSIN-MILWAUKEE

AND

DONALD M. BAER

UNIVERSITY OF KANSAS

A single-subject design often used to compare the effectiveness of two or more independent variables (like treatment programs) is the multielement (alternating treatments or simultaneous treatments) design. Variants of this design approximate the concurrent comparison of the effects of two or more variables (or levels of variables) by programming the variables (or levels) in rapid alternation, typically across or within daily sessions. Properly combined with conventional reversal designs, these designs can also display a variety of interaction effects, some of them worrisome, others highly desirable for the future development of the field. A worrisome model is the possibility that when Treatment B alternates rapidly with Treatment C, the effects of each will not be the same as when each is the only treatment used. A desirable model is the use of the multielement design as a fast-paced component of an otherwise conventional reversal design examining contextual control of some relationship: the possibility that some behavior responds differently to Controlling Variables A and B in Context X than in Context Y. This second possibility opens single-subject designs to the more efficient examination of all interactive effects and is highly desirable, considering the prevalence and importance of interactions in determining the limits and the generality of currently understood behavioral phenomena.

DESCRIPTORS: interaction effects, multiple treatment interference, multielement design, alternating treatments design, methodology

THE SPECIAL USES OF MULTIELEMENT DESIGNS

The multielement design is simply a fast-paced reversal design incorporating many reversals. Like any reversal design, it can be used to compare the effects of different levels of a variable (e.g., treatment and no treatment) or different variables (e.g., Treatment 1, Treatment 2, and Treatment 3). In the latter case, the point of using the design is often to show promptly which of those treatments is best

for subsequent prolonged use; thus, the design is usually called an alternating treatment or simultaneous treatment design (Barlow & Hayes, 1979; Barlow & Hersen, 1984; Hersen & Barlow, 1976; Kazdin, 1982; McReynolds & Kearns, 1983; Ullman & Sulzer-Azaroff, 1975). Even so, it is still a multielement design, in that this label denotes nothing more than relatively many, relatively fast-paced reversals: Its variables are changed sometimes every session and sometimes within every session.

The programming of relatively many, relatively fast-paced reversals has a highly specialized set of advantages. Perhaps the major advantage is the ability to compare variables within the context of uncontrolled and uncontrollable background variables that can be presumed to change more slowly or less often than the experimental variables will be made to change. This allows the relative effects of the experimental variables to be seen unfounded with those background variables. Thus, for example, if two teaching methods can be compared daily throughout a school term, the comparison thereby escapes confoundings with such

Preparation of this article was supported in part by the National Institute of Child Health and Human Development, National Research Service Award, Public Health Service Grant 5 T32-HD-07173, by the National Institute of Mental Health Grant R01-MH-20411-14, and by the Handicapped Children's Early Education Program Grant 024BH50009.

Correspondence and requests for reprints may be sent to Ann Higgins Hains, Department of Exceptional Education, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, or to Donald M. Baer, Department of Human Development and Family Life, University of Kansas, Lawrence, Kansas 66045.

background variables as the weather, changes in curriculum topics and difficulty, changes in teachers, extracurricular activity cycles (e.g., sports, dances, field trips), and factors vaguely hypothesized as beginning-of-year and end-of-year effects.

A secondary advantage of the multielement design is that the answer to the experimental question seems to emerge very quickly; only its generality remains in question, and if the design continues, that question begins to be answered, too, at least for those background variables that will in fact change during that time.

However, those virtues also define the design's intrinsic specialization. The multielement design requires a special kind of problem to investigate: the comparison of variables that produce immediate rather than slow and cumulative effects on the behavior under study, and that, when discontinued, allow the behavior to resume promptly its prior status. Thus, multielement designs are best suited to studying effects such as stimulus control and are ill suited to a comparison of different methods that slowly and cumulatively establish new, complex skills, and might better be examined in a slow-paced reversal design with relatively few reversals.

Within that context, a further possible advantage is less certain—that multielement designs reduce the likelihood of some types of multiple treatment interference. Multiple treatment interference is “likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable” (Campbell & Stanley, 1963, p. 6). Sequence, carry-over, and alternation effects often are mentioned as possible sources of multiple treatment interference in single-subject research (Barlow & Hayes, 1979; Barlow & Hersen, 1984; Kazdin & Hartmann, 1978; McReynolds & Kearns, 1983). These types of interaction effects are all examples of the same kinds of bias and can be addressed in the same manner.

SEQUENCE EFFECTS

Sequence effects (also known as order effects) refer to the possibility that if one treatment, say C,

is examined only after another treatment, say B, has occurred, as in an A-B-C design, then some of Treatment C's effectiveness (or ineffectiveness) may be due to the fact that B preceded it. This possibility is undeniable and extremely problematic, for by the same logic we should always ask if some of B's effectiveness (or ineffectiveness) may be due to the fact that C did not precede it. And then we should also ask whether any of B's and C's effectiveness (or ineffectiveness) may be due to the fact that A preceded them, and whether any of A's effectiveness (or ineffectiveness, especially) may be due to the fact that neither B nor C preceded it. Thus, the entire domain of designs that examine different levels or forms of variables over time within the same organism is subject to the problem of sequence effects. The proper question is whether sequence effects are indeed a problem, rather than a fact of nature that we should simply let be rather than “solve.”

When Sequence Is Considered a Problem

When sequence is considered a problem, it is often handled by counterbalancing. Proper counterbalancing simply administers A-B-C, A-C-B, B-A-C, B-C-A, C-A-B, and C-B-A sequences equally often. Occasionally that is done by assigning each sequence to a different subject. Such designs are committed to the assumption that A, B, and C interact similarly in all subjects, a very doubtful assumption. Sometimes that assumption is obviated by assigning each sequence to large enough, randomly selected groups of subjects. That obviates the entire point of single-subject designs.

Ulman and Sulzer-Azaroff argue that the multielement design “minimizes possible sequence effects by presenting each condition only briefly . . . rather than for a prolonged period of time” (Ulman & Sulzer-Azaroff, 1975, p. 387). They suggest that *within-subject* counterbalancing in the multielement designs is the appropriate form, and that it may co-opt the problem of sequence effects: If the generalized form of the problem is that B precedes C, then a useful design might well let it do so as briefly as possible, and also let C precede B just as often (and just as briefly) as B precedes C for that

subject. That is exactly a multielement design, and it is a within-subject tactic. Because multielement designs are naturally built on some repetition of sessions, it is the sessions that can meaningfully accomplish as much counterbalancing of B-C and C-B sequences as their number allows. In addition, controllable background variables such as choice of experimenters, choice of settings, and times of day also can be counterbalanced across sessions (if there are enough sessions).

Counterbalancing *does not* eliminate sequence effects; it merely allows them (or their absence) to be seen. However, when B-C sequences show different outcomes for B and C than C-B sequences do, it requires a fair number of replications of each sequence to see that as a reliable and unambiguous difference. For example, a commonly used design is to establish a baseline, A, add the treatment conditions, say B and C, to that baseline in a multielement comparison, and then use the "better" treatment as the third and final condition:

A
A-B-C.
C

This type of design and this method for diagramming it were first illustrated by Browning in 1967. Here, baseline levels of performance, A, are assessed first in isolation for several sessions as the only condition operative across the three separate periods. Following baseline, A appears in fast-paced alternation with the interventions (a typical multielement design) for several more sessions. In the multielement phase, Treatments B and C are compared every session to each other and to A, and those comparisons can be counterbalanced, requiring six kinds of sessions: A-B-C, A-C-B, B-A-C, B-C-A, C-A-B, and C-B-A. Each of these six sequences ought to be compared to the others as often as possible, if counterbalancing is to be a serious response to a serious problem; doing so implies many sessions (12 at a bare minimum, and preferably a much larger multiple of six). Then the better of B and C, say C, can be examined in isolation during the subsequent final sessions of the study. Such a design may be represented more clearly in the sessions diagrammed below:

A A A A B B C C A A B B C C C C
A-A-B-C-A-C-A-B-B-C-A-C-A-B-C-C.
A A C B C A B A C B C A B A C C

Thus, responding to potential sequence effects with counterbalancing sacrifices much of the multielement design's supposed ability to be quickly informative.

This design does not analyze sequence effects, in that it does not relate them to more fundamental variables, but it does display some of them—as the inability to recover levels of A or C (but not B) across experimental phases where they operate in alternation and in isolation. For example, changes in levels of A can be examined across baseline, where A appears in isolation (by itself across the three sessions), and in the treatment phase, where A alternates with B and C.

Do we ever fail to recover levels of behavior, either within or across experimental phases, in this design? If we do, then the multielement design has not lived up to its advertising about minimizing sequence effects. In fact, we do. A survey of the multielement designs represented in Barlow and Hersen's classic 1984 textbook revealed 31 cases, 14 of which were unambiguous for this question. Of those 14 cases, 4 cases recovered every level examined in either a multielement comparison or in isolation (Kazdin & Geesey, 1977; Shapiro, Barrett, & Ollendick, 1980; Van Houten, Nau, MacKenzie-Keating, Sameoto, & Colavecchia, 1982; Weinrott, Garrett, & Todd, 1978); however, 10 cases failed to recover prior levels across multielement and isolation comparisons (Barrett, Matson, Shapiro, & Ollendick, 1981; Bittle & Hake, 1977; Hallahan, Lloyd, Kneedler, & Marshall, 1982; Martin, Pallotta-Cornick, Johnstone, & Celso-Goyos, 1980; O'Brien, Azrin, & Henson, 1969; Ollendick, Matson, Esveldt-Dawson, & Shapiro, 1980; Ollendick, Shapiro, & Barrett, 1981; Rojahn, Mulick, McCoy, & Schroeder, 1978; Shapiro, Kazdin, & McGonigle, 1982; Singh, Winton, & Dawson, 1982). If this small sample can be taken as at least a cautionary encounter with the still unknown general case, then multielement designs, despite their excellent face characteristics, do not reliably obviate sequence effects.

If we wish to see more of those potential sequence effects, we might in principle expand the multielement design for exactly that purpose, losing even more of its value as a quickly informative design, but gaining an appreciation of what kinds of sequence effects can appear in within-subject analyses, and how often. Following the logic of the preceding example, a more complete display of sequence effects will require the following prototype:

A	A	A	A	A	A	A
A-B-B-B-C-B-A-B-B-B-C-B-A-B-	...					
C	C	C	C	C	C	C

The essence of this prototypic design is simply that every element of the design—every A, B, and C—is examined both in the characteristic fast-paced alternation of the multielement design and in isolation (and in as many sequences of that as the designer has time, curiosity, and ethical permission to pursue). This diagram assumes that when each element appears alone, it is implemented across the three sessions or time periods; likewise, it assumes that when the elements are alternated, they are counterbalanced.

CARRYOVER AND ALTERNATION EFFECTS

Carryover effects refer to “the influence of one treatment on an adjacent treatment, irrespective of overall sequencing” (Barlow & Hersen, 1984, p. 257). Alternation effects refer to multiple treatment interference resulting from the speed of alternation (rapid vs. slow) and the length of the intercomponent interval separating treatment conditions (Barlow & Hersen, 1984; McGonigle, Rojahn, Dixon, & Strain, 1987). Sequence, carryover, and alternation effects are usually discussed separately; however, they all refer to the same problem—“that one experimental treatment is interfering with the other within the experiment itself” (Barlow & Hersen, 1984, p. 257). Carryover effects are nothing but sequence effects of one component of a multielement phase preceding or following other components *within that phase*. The prototypic design above shows the possibility of both within-the-

multielement-phase sequence effects and between-phases sequence effects by comparing the levels of any A, B, or C both within and between all its phases, multielement and in-isolation phases alike. Thus, there is little reason to maintain a distinction in terminology between sequence, carryover, and alternation effects. All that is at issue are sequence effects, sometimes in faster paced sequences, sometimes in slower paced sequences.

If within-the-multielement-phase sequence effects are present—if, for example, C looks as effective as it does only because it alternates so quickly with B—they will become apparent when C is examined in isolation from B, *if* C is examined at sufficient length when it is separated from B. If C looks as effective as it does in the alternation phase with A and B only because B is present, perhaps it can continue to look that effective for a few more sessions when it is examined in isolation in the next condition of the design. Then, that next condition of C-only must be a protracted one. Otherwise, these designs, instead of building a technology for improving deviant behavior, will yield a classic literature of contradictory results, specifically one in which later use of Treatment C does not match the promise shown in its initial “validation” research.

To deal with all these problems of potential sequence effects, slow- or fast-paced, the prototypic design needs even more elaboration:

A	A	A
A-B-C-B-AAAAA-B-BBBBBB-B-CCCCC-	...	
C	C	C

This design is a very lengthy one; indeed, it is so lengthy that in the pragmatic context of comparing two clinical treatments, it is unlikely ever to be carried out and probably should not be. This is especially so because of the obvious ethical considerations of (a) not treating subjects needful of a prompt, optimum intervention and (b) treating them as ping-pong balls, to be eternally bounced back and forth between conditions, sometimes slowly, sometimes quickly, for the sake of a complete examination of treatment–sequence interactions.

When Sequence Is Not Considered a Problem

In the pragmatic context of treatment, the Browning (1967) design is good enough, if its subsequent examination of the better treatment, say C, is protracted enough, and especially if it seems successful enough to allow its eventual end:

$$\begin{array}{c} A \\ A-B-C \dots -A \dots \\ C \end{array}$$

The enabling arguments are essentially pragmatic ones. If C looks as effective as it does during the multielement phase, in part because of its fast-paced alternation with B, then that should become apparent in the subsequent protracted examination of C alone: C alone eventually will look different from C alternating rapidly with A and B. In particular, if C alone eventually seems more effective than it had during the multielement phase, we may reasonably suspect that its fast-paced alternation with A and B diluted its effectiveness then, but we will hardly care, because the design gave us the correct answer for the next clinical phase of our program. On the other hand, if C seems to lose the effectiveness that it had shown during the multielement phase, that will set the occasion for a subsequent examination of B alone, to see if it will be more effective, now that we suspect that its earlier inferiority to C was illusory (i.e., was an artifact of the multielement design).

This finding can suggest that B was always the superior treatment, but looked inferior to C only because of its fast-paced alternation with A and C. However, if B is that fragile a superior treatment—if it can become that ineffective simply by its proximity to some A or C—then perhaps that recommends against relying on it in a treatment context: How can we ensure that the uncontrolled part of our client's everyday life will not present some deleterious partner to our B in the course of treatment? In other words, we probably do not want to use Bs that sensitive to uncontrollable events.

If A alone at the end of the study seems much more effective than A alone as baseline or during the multielement phase of the design, that will simply testify to an enduring effect of treatment.

We may well begin asking what natural community of support for the target behavior has been tapped by the treatment, a question that will require quite a different investigation rather than a better version of the above design.

In short, in a treatment context, we are usually not interested in the complete analysis of our target behavior's responsiveness to the sequences of the controlling variables that we can apply. Instead, we seek variables that produce useful effects *despite* their sequencing with other variables (useful or otherwise). Everyday life, as far as we know, is an ongoing, unassessable set of such sequences, some of them segregated in their effects to the settings in which they operate, some of them showing generalized effects across an unpredictable range of settings. But we can eventually restrict our focus to those variables that operate uniformly despite such sequences by examining them in reversal designs that make clear their durability or their fragility in such sequences. However, doing so will still require an ability to examine those sequence-durable effects across contexts more fundamental than the order of presentation, and for that purpose, the fast pace of the multielement design may have special usefulness, as the next section will show.

Other Approaches to the Problem

The procedures Sidman (1960) termed *independent verification* and *functional manipulation* are variations of the previously discussed prototypic designs. Sidman argues that these procedures assess the extent of any fast-paced interaction effects but do not control or prevent them. Independent verification enables the experimenter to study the "simultaneous control of behavior by a multiplicity of variables" (Sidman, 1960, p. 335) by programming each variable of interest both alone and in combination with the other variables of interest (Barlow & Hersen, 1984; McReynolds & Kearns, 1983; Shapiro et al., 1982; Sidman, 1960). A narrower definition of an independent-verification interaction design is that it follows the traditional slower paced A-B-A-B reversal design, but arranges conditions to display the ongoing additive or sub-

tractive effects of each variable (McReynolds & Kearns, 1983); sequence effects are of course possible, but will be seen as such in a long enough design by examining the variables across conditions. Variables may appear in combination, BC, or in alternation, $\overset{B}{C}$. For example, verbal and nonverbal reinforcement could be scheduled simultaneously or in alternation. The design sometimes adds a variable:

$$A-B-BC-B-BC \quad \text{and} \quad A-B-\overset{B}{C}-B-\overset{B}{C}$$

and sometimes subtracts a variable:

$$A-BC-B-BC-B \quad \text{and} \quad A-\overset{B}{C}-B-\overset{B}{C}-B.$$

When two variables are introduced at the same time in a package (e.g., BC), a concurrent interaction is always possible, as Sidman recognizes. For example, in an analysis of the effects of adaptive clothing on the self-injurious behavior of two blind, profoundly retarded men (Rojahn et al., 1978), a jacket (B) and a neck-brace (C) were examined separately, as a package (BC), and as fast-paced alternate conditions within a multielement phase:

$$A-B-BC-\overset{A}{BC}-\overset{B}{C}-BC.$$

In this design, there is probably as much possibility of interactive effects in the packaging of B and C as there is in their fast-paced alternation in the multielement comparison. The display of interaction effects is limited in this design, however, because the design does not systematically replicate isolation phases with either the packaging of BC or the alternation of $\overset{B}{C}$. Yet, the approximation of the prototypic design seen above was an appropriate beginning and would have been even if there had not been a multielement phase.

A more complex examination of interaction effects through independent verification is seen in a study by Shapiro et al. (1982). This study was designed to examine treatment interactions within a multielement design. In addition to examining time-of-day effects (morning vs. afternoon), the design sometimes used fast-paced alternations of baseline (A), token reinforcement (B), and token

reinforcement with response-cost contingencies (BC) on the on-task behavior of mentally retarded, behaviorally disturbed children:

$$A-\overset{A}{B}-\overset{B}{BC}-\overset{A}{B}-\overset{B}{BC}-B.$$

This design does not examine response-cost effects, C, in isolation; rather, it only compares response cost as a packaged treatment, BC, with the token reinforcement element, B. The examination of the response-cost system in isolation would not have been possible without some existing token system. Token reinforcement, however, was compared in alternation with baseline, $\overset{A}{B}$, and then in alternation with the token reinforcement and response cost contingencies, $\overset{B}{BC}$. The resultant data showed the effects of several interactions: On-task behavior during the token reinforcement condition (B) was more variable when that condition alternated rapidly with the token reinforcement and response-cost condition ($\overset{B}{BC}$) than when it alternated with the baseline condition ($\overset{A}{B}$). Probable sequence effects also appeared: Comparison of the initial $\overset{A}{B}$ (fast-paced alternation of baseline and token reinforcement conditions) and its replication showed improved on-task behavior in both conditions in the replication. Furthermore, the introduction of the token reinforcement and response-cost phase apparently influenced subsequent performance in both token and baseline conditions.

An examination of the studies cited in the Barlow and Hersen (1984) text revealed that only three studies used designs that allowed some examination of interactions of treatment variables through independent verification (Bittle & Hake, 1982, and the two studies previously described). Each of these studies used an approximation to one of the prototypic independent-verification designs. In general, whenever variables appear in fast-paced alternation ($\overset{B}{C}$) or in packages (BC) (which might be considered the ultimate case of fast pacing), their repetitive examination in isolation as well is worth considering, especially if we wish to use less ambiguous terms than the "apparently" and "probably" necessary in the above descriptions of these results. The consideration of an independent-verification design

should be subject to the standard benefit-cost logic of applied research, of course. There may well be times when the benefit of being less ambiguous about those possibilities will not seem worth the cost. Furthermore, there may be situations in which the treatments compared in a multielement design could not be combined because the variables are procedurally incompatible. Similarly, situations may exist where components of a package intervention could not be tested separately.

Sidman illustrates a second method called *functional manipulation* (1960, p. 336). By systematically altering some parameter of an experimental variable, the researcher can see if those changes affect the relationship between that variable and the behavior under study.

Two studies cited by Barlow and Hersen (1984) provide examples of investigations in which important variables were examined in interaction (Corte, Wolf, & Locke, 1971; Doke & Risley, 1972). The Corte et al. (1971) study cited by Barlow and Hersen provides an elegant example of what could well be termed a superordinate multielement analysis, wherein two variables were examined in interaction, each in its own multielement design, one design within the other. That study examined the effects of food deprivation and a DRO using food reinforcement on the self-injurious behavior of an institutionalized profoundly retarded adolescent. The subject's lunch was withheld every other day, so that deprivation, D, and nondeprivation, d, conditions alternated with each daily session. Brief periods of contingent reinforcement of other behavior, R, and of noncontingent reinforcement, r, both with bites of food, occurred within each daily session; the contingent condition always preceded the noncontingent condition:

$$D^R-d^R-D^r-d^r \dots$$

The results showed that the lowest rates of self-injurious behavior occurred under deprivation and contingent reinforcement conditions. The highest rates of self-injurious behavior occurred during the nondeprivation and noncontingent reinforcement conditions.

The design of the Corte et al. (1971) study is a multielement design within another, slower paced multielement design. That is, contingent and noncontingent reinforcement conditions alternated within each session, and deprivation and nondeprivation conditions alternated with each successive session. That design permitted the detection of the "better" treatment procedure for eliminating self-injurious behavior, but, more important for the future of the discipline, it examined that superiority within the context of an important parameter of both treatments: The meaning of reinforcement is such that it cannot be examined except at some point on its deprivation-satiation interaction, and the meaning of contingency is such that it cannot be examined except in some contrast to a somewhat different contingency. Because these variables *must* interact, they are better studied as interactions than not, and the design just described may well be the simplest one for doing so.

If the design were extended, it might also allow an inspection of the effects of contingent and noncontingent reinforcement in isolation from each other, meaning of course only in a slower paced alternation (which is, logic suggests, all that "isolation" can mean in single-subject designs and in real life):

$$D^R-d^R-D^r-d^r-DR-dR-Dr-dr \dots$$

This design might offer the best of both worlds: the same relatively quick display of an interaction that the prior design offered, and an only slightly slower display of any sequence effects that might be operating within that interaction. Again, the design is obviously a costly one, and benefit-cost logic applies, as ever.

Doke and Risley (1972) provide another example of superordinate multielement analysis. This study examined the effects on preschool children's participation in activities during group versus individual dismissal from the activities; these activities were sometimes with materials (e.g., house-keeping) and sometimes without (e.g., story time). The conditions of activities with materials, M, and without materials, m, alternated within each day,

always at the same times of day. The dismissal conditions alternated every few days; for a few days, children were dismissed from all of their activities as a group (as soon as all children were ready), G; then, for a few more days, they were dismissed individually (as soon as each child finished), g:

$$G_m^M - g_m^M - G_m^M - g_m^M \dots$$

The results showed that the highest rate of participation occurred when there were materials and dismissal was individualized.

This design also permits the detection of the "better" treatment for promoting children's participation in activities; more important, it examined that superiority within the context of an important parameter, materials. All play exists at some point on a dimension of material use and availability; and all adult-managed play exists on some dimension of termination of or dismissal from that play. Thus, the study of play *ought* to be the study of at least these interactions, and again, the above design may represent the minimum design competent to begin that study. In this respect, the study differs from the Corte et al. (1971) study only in that some of the parameters for further manipulation may not be as apparent. For example, there are parameters underlying the effectiveness of activities without materials (e.g., stories) for children's participation: time of day, length of session, number of teachers present, and the complex of parameters called teacher "style." On the other hand, activities with materials (housekeeping, block, manipulative, and creative activities) are almost surely sensitive to somewhat different underlying parameters: the amount of materials available per child, the space available per child, and the range of playmates available. Thus, the design could be extended to examine the effects of individual and group dismissal within the context of availability of materials (many, 1; few, 2) for only activities with materials, M:

$$G_m^M - g_m^M - G_m^M - g_m^M - GM_1^1 - gM_1^1 - GM_2^1 - gM_2^1$$

Then the design could be extended to allow an inspection of the effects of many and few materials in isolation from each other:

$$G_m^M - g_m^M - GM_1^1 - gM_1^1 - GM_1^1 - gM_1^1 - GM_2^1 - gM_2^1$$

This slower extension could display any sequence effects that might be present within the interaction of the previous phase.

In the past, single-subject designs have been used in their relatively simple forms and thereby have revealed to investigators a wealth of powerful direct effects in the analysis of behavior. By the same token, they have obviated the study of contextual factors (such as deprivation, availability of materials, task difficulty) in those effects: We have learned that certain processes are very powerful, but usually have failed to learn the contextual conditions that maximize and minimize that power. Yet almost certainly, every process fundamental to the analysis of behavior is subject to exactly that kind of contextual control (cf. Kantor, 1959; Morris, Higgins, & Bickel, 1982); and almost certainly, contextual control will be found to be very powerful in modulating the generality of those apparently powerful processes. If so, it is not the final details of the analysis of behavior that have escaped investigation in the simple single-subject designs of our history; it is the fundamental statements of generality and the fundamental conditions of a *dependable* technology that await clarification.

The research paradigm for that clarification must involve a study of interactions. In its simplest form, the interaction displays the effects of several levels of Variable 1 in the context of several levels of Variable 2. For example, the effects of peer prompting of a child's social behavior may be different when certain peers do the prompting than when other peers do it, and within each of those classes, the effects may be different depending on the peer's rate of prompting. Suppose that a useful kind of peer status can be defined in terms of the subject's typical past rate of playing with that peer. Do peer status and prompt rate interact? In other words, does peer status determine the effectiveness of peer-prompt rate? Or, the same question with a somewhat different theoretical implication, does prompt rate determine the effectiveness of the prompting peer's status? The appropriate design will examine four conditions, at least (see Figure 1).

The familiar four-fold table as shown in Figure 1 has been the symbol of group-factorial designs for many years in behavioral research; in that tradition, each cell of the table has usually represented another group of subjects. Perhaps the most significant aspect of the multielement design is that it allows us to consider the same interaction (indeed, a more realistic version of it) within a single subject. If *s* and *S* represent low-status and high-status peers, respectively, and *r* and *R* represent low rates and high rates of their prompts, respectively, the following multielement within multielement designs will examine the question:

$$R_s^s - R_s^R - r_s^s - r_s^R - R_S^s - R_S^R - r_S^s - r_S^R$$

and

$$s_r^R - s_r^s - S_r^R - S_r^s - s_R^R - s_R^s - S_R^R - S_R^s$$

The difference in these designs is primarily the difference implied in the above figure by labeling one variable TREATMENT and the other PARAMETER. That difference represents some difference in theory, no doubt; more important is the question of whether the data of the two designs would yield the same answer or two somewhat different ones.

Next, consider extending these designs to include the meaningful components examined in isolation as well. That allows the same examination of the interaction of these variables—the effects of prompt rates in the context of peer status (or vice versa)—as could be considered in the group design, and also allows an examination of whether sequence effects will operate as a consequence of examining this interaction within a single subject, and if so, to what extent.

Perhaps the most interesting case that might result from this examination is if sequence effects do appear. Some research methodologists will then argue that the group-factorial design is clearly the superior one, in that it does not allow those sequence effects to operate, and thus allows *the* interaction between rates of prompting social behavior and peer status to appear in its purest form. The point, however, is to recall Sidman's (1960) argument about such pure forms of behavioral phenomena: If they do not operate that way within the single

		TREATMENT	
		High Rate Prompts	Low Rate Prompts
PARAMETER	High Status Peers		
	Low Status Peers		

Figure 1. Examination of interactions between high-status and low-status peers and high-rate and low-rate prompts.

subject whose behavior we are trying to analyze, then no matter what their purity when freed of sequence effects through the segregation of their component variables to different subjects, they are not the analysis of *this* subject's behavior (and, indeed, perhaps not the analysis of *any* subject's behavior).

Students may encounter rapid alternations of high-status and low-status peers in their social environment, and those peers may often vary their rates of social bids. The real-world question is how those contextual parameters affect social behavior under *those* conditions of constant encounters with all levels of the variables under investigation. If the answer is that sequence effects operate within that interaction, that is the *realistic* answer to our curiosity, and we should simply let the design continue until all such effects (differences and nondifferences) seem stable. After all, that is close to what happens to students in the real world, and so that answer ought to have some real-world generality. The results of the group-factorial design, precisely because it has been purged of that sequence effect, give a purely unrealistic picture of the interaction between peer status and rate of prompting social behaviors.

These parametric manipulations are meaningful not only because they detect interactions present in a given set of data, but also because they allow the researcher to create or eliminate selected interactions through functional manipulation. In other words, when these parameters are amenable to precise enough experimental control and that control is exercised, the result is a truly systematic display of the interaction in question. For example, in the design of the Corte et al. (1971) study, we might well go on to examine various levels of deprivation,

		TREATMENT	
		Posted	Not Posted
PARAMETER	High Density		
	Low Density		

Figure 2. Examination of interactions between posted and nonposted speeds and high and low densities of police cars.

various schedules of reinforcement, and, within each of those schedules, various levels of the schedule's parameters (e.g., ratio size or interval length); knowing all those effects on self-injurious behavior might indeed give us a precise treatment technology for that problem.

Similarly, parametric manipulations could detect interactions in more applied problems. Consider two hypothetical examples. First, the speeding behavior of motorists could be examined in a multielement design wherein on some days the speed at which motorists were traveling was posted for their observation, and on other days it was not. In addition, the visible density of police cars in the area could be varied to be high, say 20 units, and low, 2 units, on a weekly basis as shown in Figure 2. (Again, which is TREATMENT and which is PARAMETER is optional.) By deliberately altering one of the components—say, the schedule of posting—changes in, or the stability of, the relationship of speeding to the other variables could be examined.

Van Houten et al. (1982) offer a second example in two separate studies. In one study, they used a multielement design to compare the effects of verbal reprimands versus a package of verbal reprimands, eye contact, and a firm grasp on the student's shoulder on student disruptive behavior. In a second study of student disruptive behavior, they compared the effects of reprimands without eye contact when the teacher was close to the student (1 m away) with reprimands given from a greater distance (7 m). Figure 3 shows how these two studies could be combined to examine the interactive effects of some of these variables. This four-fold table can

		TREATMENT	
		Reprimands	
		Verbal Only	Verbal & Eye Contact
PARAMETER	1 Meter Away		
	7 Meters Away		

Figure 3. Examination of interactions between verbal reprimands alone and with eye contact at distances of 1 m and 7 m.

be realized in multielement designs, just as were the previous examples. Perhaps any four-fold table of experimentally manipulable variables can; we need to find out. But more urgently, we need to reconsider what variables are worth the appreciable cost of these designs.

Presently, many studies using multielement designs have examined the parameters of time periods (morning vs. afternoon), settings (classroom vs. therapy session), or adults (male vs. female). These interactions are usually uninteresting; they lack both generalizability and explanation. In other words, when interactions are found between treatments and such "marker" variables (Baer, 1984), the results are nonexplanatory: If morning sessions produce better treatment effects than afternoon sessions, numerous speculations could be made regarding why this effect occurred. Sometimes the reason is eventually identified—say, differential fatigue produces the differences between sessions. Yet, even the knowledge about this factor need not identify the fundamental processes responsible for the interaction effect (why should fatigue interact with this treatment in this way?). And it need not specify the nature of its further investigation (although its management may be clinically desirable even in the absence of understanding why it occurs).

Extraneous variables can be and often are counterbalanced in multielement designs, so that their influence on different treatments will be balanced—they will have equal opportunity to influence each treatment's effectiveness. Several authors recommend additional procedures for minimizing multiple treatment interference, such as separating

treatment sessions with a time interval, and using slower and/or presumably more discriminable alternations (Barlow & Hersen, 1984; Kazdin, 1982; McGonigle et al., 1987; McReynolds & Kearns, 1983). These authors carefully assert that these procedures only *minimize* multiple treatment interference (and the more prudent of these authors assert even more carefully that these procedures only *probably* minimize those effects). Even so, some studies continue to claim that they have controlled for interaction effects by using these techniques. Sequence effects, slow- or fast-paced, may or may not be eliminated by these techniques, and certainly could be present in any fast-paced alternation of treatments, counterbalanced or not. These procedures may be useful in minimizing such effects, but they do not control for them either in the sense of providing experimental evidence of when these interactions occur, or in the sense of preventing them from occurring. Further experimental manipulations are necessary to assess directly the extent to which interaction effects are present in multielement designs.

Even when precise control of marker variables is possible, the nature of their interactions quite often is specific to the individual subject (e.g., McGonigle et al., 1987). The magnitude and intricacies of such interactions usually will be idiosyncratic (i.e., not generalizable to a wide variety of conditions or subjects). Consequently, the value of experimentally investigating behavior under the interactive control of these variables will be correspondingly restricted. For these reasons, researchers should be encouraged to analyze interactions between treatments and more meaningful contextual variables rather than to counterbalance them.

Unfortunately, few researchers have been encouraged to do so. Most of the multielement designs reviewed by Barlow and Hersen (1984) did not examine interactions among variables and were not intended to do so. Only 14 of their 31 reviewed studies used multielement designs that allow some examination of interaction effects, and of these, only two—those of Shapiro et al. (1982) and Van Houten et al. (1982)—explicitly proposed to do so. The other 12 studies did not mention the extent

to which interaction effects were present, and they did not evaluate the usefulness of their designs for examining interactions.

A survey of the 1975 and 1984 volumes of the *Journal of Applied Behavior Analysis* also supports the conclusion that the analytic use of multielement designs has not changed. Fourteen studies used multielement designs in the 1984 volume, whereas only seven appeared in 1975; but neither volume included studies that examined interaction effects.

CONCLUSION

When multielement designs are set within a reversal design alternating fast-paced alternations of their components with examinations of each element in isolation, as shown in the prototypic design variations above, they are capable of revealing when sequence effects operate, as described earlier by Sidman in his sketch of the methods of independent verification and functional manipulation (1960). Unfortunately, the authors of single-subject research design books usually discuss these methods only as assessing multiple treatment interference. Any interpretation of that literature as “employ the appropriate procedural controls (such as counterbalancing) and hope to minimize the effects of multiple treatment interference” seems misguided; instead, “assess many effects of potential multiple treatment interference” and “try superordinate multielement designs for the study of contextual interactions more meaningful than sequence effects” both seem more accurate and realistic. So far, researchers have often investigated the relative effectiveness of two or more treatments in alternation, but they have rarely examined whether the treatments interact, and they have not investigated the generality of their treatments’ effectiveness by varying their potentially crucial contextual parameters (cf. Van Houten, 1987).

Ironically, the multielement design was introduced initially as a quickly informative design. Within very severe limits, it is. But the experience in the field of using it for that purpose now shows us that it can have a much more valuable func-

tion—the study of interactions. To serve that function, the multielement design will prove invaluable, but now it will be an expensive design: It will cost a great deal of time and careful experimental control over many conditions. In the world of experimental design, perhaps we should always doubt that a great deal of information can ever be gained in a very short design.

REFERENCES

- Baer, D. M. (1984). Future directions: Or is it useful to ask, "Where did we go wrong?" before we go? In R. F. Dangel & R. A. Polster (Eds.), *Parent training: Where it's at and where it's going* (pp. 547–557). New York: Guilford Press.
- Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis*, **12**, 199–210.
- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.
- Barrett, R. P., Matson, J. L., Shapiro, E. S., & Ollendick, T. H. (1981). A comparison of punishment and DRO procedures for treating stereotypic behavior of mentally retarded children. *Applied Research in Mental Retardation*, **2**, 247–256.
- Bittle, R., & Hake, D. F. (1977). A multielement design model for component analysis and cross-setting assessment of a treatment package. *Behavior Therapy*, **8**, 906–914.
- Browning, R. M. (1967). A same-subject design for simultaneous comparison of three reinforcement contingencies. *Behaviour Research and Therapy*, **5**, 237–243.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Corte, H. E., Wolf, M. M., & Locke, B. J. (1971). A comparison of procedures for eliminating self-injurious behavior of retarded adolescents. *Journal of Applied Behavior Analysis*, **4**, 201–213.
- Doke, L. A., & Risley, T. R. (1972). The organization of day-care environments: Required vs. optional activities. *Journal of Applied Behavior Analysis*, **5**, 405–420.
- Hallahan, D. P., Lloyd, J. W., Kneedler, R. D., & Marshall, K. J. (1982). A comparison of the effects of self- versus teacher-assessment of on-task behavior. *Behavior Therapy*, **13**, 715–723.
- Hersen, M., & Barlow, D. H. (1976). *Single-case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press.
- Kantor, J. R. (1959). *Interbehavioral psychology*. Granville, OH: Principia Press.
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kazdin, A. E., & Geesey, S. (1977). Simultaneous-treatment design comparisons of the effects of earning reinforcers for one's peers versus for oneself. *Behavior Therapy*, **8**, 682–693.
- Kazdin, A. E., & Hartmann, D. P. (1978). The simultaneous-treatment design. *Behavior Therapy*, **9**, 912–922.
- Martin, G., Pallotta-Cornick, A., Johnstone, G., & Celso-Goyos, A. (1980). A supervisory strategy to improve work performance for lower functioning retarded clients in a sheltered workshop. *Journal of Applied Behavior Analysis*, **13**, 183–190.
- McGonigle, J. J., Rojahn, J., Dixon, J., & Strain, P. S. (1987). Multiple treatment interference in the alternating treatments design as a function of the intercomponent interval length. *Journal of Applied Behavior Analysis*, **20**, 171–178.
- McReynolds, L. V., & Kearns, K. P. (1983). *Single-subject experimental designs in communicative disorders*. Baltimore: University Park Press.
- Morris, E. K., Higgins, S. T., & Bickel, W. K. (1982). Comments on cognitive science in the experimental analysis of behavior. *The Behavior Analyst*, **5**, 285–290.
- O'Brien, F., Azrin, N. H., & Henson, K. (1969). Increased communications of chronic mental patients by reinforcement and by response priming. *Journal of Applied Behavior Analysis*, **2**, 23–29.
- Ollendick, T. H., Matson, J. L., Esveldt-Dawson, K., & Shapiro, E. S. (1980). Increasing spelling achievement: An analysis of treatment procedures utilizing an alternating treatments design. *Journal of Applied Behavior Analysis*, **13**, 645–654.
- Ollendick, T. H., Shapiro, E. S., & Barrett, R. P. (1981). Reducing stereotypic behaviors: An analysis of treatment procedures utilizing an alternating treatments design. *Behavior Therapy*, **12**, 570–577.
- Rojahn, J., Mulick, J. A., McCoy, D., & Schroeder, S. R. (1978). Setting effects, adaptive clothing, and the modification of head-banging and self-restraint in two profoundly retarded adults. *Behavioural Analysis and Modification*, **2**, 185–196.
- Shapiro, E. S., Barrett, R. P., & Ollendick, T. H. (1980). A comparison of physical restraint and positive practice overcorrection in treating stereotypic behavior. *Behavior Therapy*, **11**, 227–233.
- Shapiro, E. S., Kazdin, A. E., & McGonigle, J. J. (1982). Multiple-treatment interference in the simultaneous- or alternating-treatments design. *Behavioral Assessment*, **4**, 105–115.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books.
- Singh, N. N., Winton, A. S., & Dawson, M. H. (1982). Suppression of antisocial behavior by facial screening using multiple baseline and alternating treatments designs. *Behavior Therapy*, **13**, 511–520.

- Ulman, J. D., & Sulzer-Azaroff, B. (1975). Multielement baseline design in educational research. In E. Ramp & G. Semb (Eds.), *Behavior analysis: Areas of research and application* (pp. 377-391). Englewood Cliffs, NJ: Prentice-Hall.
- Van Houten, R. (1987). Comparing treatment techniques: A cautionary note. *Journal of Applied Behavior Analysis*, **20**, 109-110.
- Van Houten, R., Nau, P. A., MacKenzie-Keating, S. E., Sameoto, D., & Colavecchia, B. (1982). An analysis of some variables influencing the effectiveness of reprimands. *Journal of Applied Behavior Analysis*, **15**, 65-83.
- Weinrott, M. R., Garrett, B., & Todd, N. (1978). The influence of observer presence on classroom behavior. *Behavior Therapy*, **9**, 900-911.

Received October 26, 1987

Initial editorial decision January 23, 1988

Revision received April 13, 1988

Final acceptance July 23, 1988

Action Editor, R. Wayne Fuqua